Statistics 581, Autumn Quarter 2007
Problem Set 2

Reading: Ferguson, Sections 1–3.

Problem 5 (Theorem of the Unconscious Statistician, 4 points).

(a) Apply the simple function principle to prove the Theorem of the Unconscious Statistician. Specifically, suppose that $(\Omega, \mathcal{A}, P)$ is a probability space, X is a random variable on $(\Omega, \mathcal{A})$ and g is a measurable function from $(\bar{\mathbb{R}}, \bar{\mathcal{B}})$ to $(\bar{\mathbb{R}}, \bar{\mathcal{B}})$. Show that

$$E_P(g(X)) = \int_\Omega g(X(\omega))\, \mathrm{d}P(\omega) = \int_{\bar{\mathbb{R}}} g(x)\, \mathrm{d}P_X(x)$$

in the sense that whenever one of the expectations exists then the other exists, and their values are equal.

(b) Give a simple, but interesting example that illustrates the conceptual difference between the two expectations above.

Problem 6 (characterization of mean and median, 4 points).

(a) If X is an integrable random variable with mean $\mu = E(X)$, show that the function

$$f(a) = E\left((X-a)^2 - X^2\right)$$

has a minimum at $a = \mu$.

(b) If X is a random variable on a probability space $(\Omega, \mathcal{A}, P)$, any number m for which

$$P(X \le m) \ge \frac{1}{2} \qquad \text{and} \qquad P(X \ge m) \ge \frac{1}{2}$$

holds is a *median* of X. Show that if m is a median of X then the function

$$f(a) = E\left(|X-a| - |X|\right)$$

has a minimum at $a = m$.

(c) Discuss the formulation of the minimization problems in parts (a) and (b). What is the scope of the subtraction?

Problem 7 (inequalities for tail probabilities, 6 points). Let X be a random variable with mean μ and variance σ^2.

(a) Prove the *Chebyshev inequality*,

$$\operatorname{pr}\left(|X-\mu| \geq r\right) \leq \frac{\sigma^2}{r^2}, \qquad r > 0. \tag{1}$$

(b) Suppose now that the distribution of X is unimodal about a mode ν, that is, X permits a Lebesgue density f which is nondecreasing up to ν and nonincreasing thereafter. Let $\tau^2 = E\,(X-\nu)^2$ be the expected squared deviation from this mode. Prove the *Gauss inequality*,

$$\operatorname{pr}\left(|X-\nu| \geq r\right) \leq \begin{cases} 1 - \dfrac{1}{\sqrt{3}}\dfrac{r}{\tau}, & r \leq \dfrac{2}{\sqrt{3}}\,\tau, \\[2ex] \dfrac{4}{9}\dfrac{\tau^2}{r^2}, & r \geq \dfrac{2}{\sqrt{3}}\,\tau. \end{cases} \tag{2}$$

Hint: Proceed in the following steps.

(i) Consider the random variable $Z = |X-\nu|$. Show that Z has Lebesgue density $g(z) = f(\nu+z) + f(\nu-z)$ which is nonincreasing for $z > 0$. Define $G(x) = \int_0^x g(z)\,\mathrm{d}z$ for $x > 0$. Without loss of generality, we may assume that $g(r) > 0$ and that $\tau^2 = \int_0^\infty z^2\,g(z)\,\mathrm{d}z$ is finite.

(ii) Pick $s > r$ such that $\operatorname{pr}\left(|X-\nu| \geq r\right) = g(r)\,(s-r)$ and show that

$$\begin{aligned} g(r)\,(s-r) &\leq \frac{4}{9r^2}\,g(r)\int_0^s z^2\,\mathrm{d}z \\ &= \frac{4}{9r^2}\left(g(r)\int_0^r z^2\,\mathrm{d}z + \int_r^s z^2\left(g(r)-g(z)\right)\mathrm{d}z + \int_r^s z^2\,g(z)\,\mathrm{d}z\right). \end{aligned}$$

Find bounds on the right hand side to show that

$$\operatorname{pr}\left(|X-\nu| \geq r\right) = 1 - G(r) \leq \frac{4}{9}\frac{\tau^2}{r^2}, \qquad r > 0.$$

(iii) Show that $1 - G(r)$ is convex for $r > 0$. Use this fact, the latter inequality and an analytic or geometric argument to prove the first part of the Gauss inequality.

(c) Is the Gauss inequality sharp? If yes, give an example of a unimodal random variable X with mode ν that attains the Gauss bound (2) for some $r > 0$. If no, what part of the proof might allow for a sharper estimate?

(d) Suppose again that the distribution of X is unimodal about a mode ν. Let $\alpha \in \mathbb{R}$ and let $\rho^2 = E\,(X - \alpha)^2$ denote the respective expected squared deviation. Vysochanskiĭ and Petunin (1980) proved that

$$\mathrm{pr}\,(|X - \alpha| \geq r) \leq \begin{cases} \dfrac{4}{3}\dfrac{\rho^2}{r^2} - \dfrac{1}{3}, & r \leq \dfrac{2\sqrt{2}}{\sqrt{3}}\,\rho. \\ \dfrac{4}{9}\dfrac{\rho^2}{r^2}, & r \geq \dfrac{2\sqrt{2}}{\sqrt{3}}\,\rho. \end{cases} \tag{3}$$

In particular, if $\alpha = \mu$ and $r = 3\sigma$ then (3) reduces to the *three sigma rule*,

$$\mathrm{pr}\,(|X - \mu| \geq 3\sigma) \leq \frac{4}{81} < 0.05. \tag{4}$$

Illustrate and compare the Chebyshev, Gauss and Vysochanskiĭ-Petunin inequalities by plotting the respective bound versus r/ρ.

Problem 8 (Chebyshev's other inequality, 2 points). Let X be a random variable taking values on a finite or infinite interval I. Let f and g be monotonically increasing real-valued functions on I.

(a) Show that

$$E f(X) g(X) \geq E f(X)\, E g(X),$$

provided that all expectations exist.

Hint: Consider $Z = (f(X) - f(X'))(g(X) - g(X'))$ where X and X' are independent identically distributed random variables.

(b) Use the result in part (a) to show that $\mathrm{cov}(f(X), g(X)) \geq 0$, provided that $E f(X)^2$ and $E g(X)^2$ are finite.

(c) Suppose now that X takes positive values only and $E X^\alpha$ is finite for all $\alpha \in \mathbb{R}$. Use the result in part (a) to prove inequalities between $E X^\alpha$, $E X^\beta$ and $E X^{\alpha+\beta}$ where $\alpha, \beta \in \mathbb{R}$. Use the result in part (b) to determine the sign of $\mathrm{cov}(X^\alpha, X^\beta)$ for $\alpha, \beta \in \mathbb{R}$.

Tilmann Gneiting, October 5, 2007. Solutions are due Friday, October 12 at the beginning of the class session.